WHAT IS CLAIMED IS:

1. A method of encoding input data within a system, wherein the input data might include sequences of symbols that repeat in the input data or occur in other input data encoded in the system, the method comprising:

identifying, within a number of sequential input data symbols defined by an offset and a window size, a fingerprint representation of the number of sequential input data symbols;

determining, from the fingerprint representation, whether the offset is to be designated as a cut point;

repeating the above steps of identifying and determining to arrive at a set of cut points;

segmenting the input data as indicated by the set of cut points;

for each segment, determining whether the segment is to be a referenced segment or an unreferenced segment;

for each referenced segment, replacing the segment data of the referenced segment with a reference label;

for each referenced segment not already present in a persistent segment store, storing a reference binding in the persistent segment store, wherein a reference binding associates a referenced segment's data and its reference label;

determining whether any sequence of segments is to be grouped as a reference group;

for each reference group, replacing the references in the group with a group label; and

for each reference group not already present in the persistent segment store, storing a group reference binding in the persistent segment store, wherein a group reference binding associates a reference group's references with its group label.

2. The method of claim 1, further comprising:

recursively identifying groups of labels into higher level groups, wherein groups of labels are one or more of groups of reference labels and groups of group labels;

for each higher level group, replacing the higher level group with a group label; and

for each higher level group not already present in the persistent segment store, storing a group reference binding in the persistent segment store for the higher level group.

3. The method of claim 1, wherein the input data comprises payloads of messages between clients and servers in a client-server network.

4. The method of claim 1, wherein the input data comprises portions of files in an on-line backup system, further comprising representing files in the on-line backup system as sequences of at least one of reference labels and group labels, and storing contents of the persistent segment store as part of the on-line backup system.

5. The method of claim 1, wherein the input data comprises portions of files in a file system, further comprising representing files in the file system as sequences of at least one of reference labels and group labels and a segment store.

6. The method of claim 1, wherein the input data comprises portions of files to be used in a file system, the method further comprising:

when storing a file to the file system, encoding it with at least one segment of the file being represented as a segment referenced in the persistent segment store; and

when retrieving a file from the file system, caching the file in a local file store as a decoded file, wherein each reference label and each group label is replaced with corresponding segment data from the persistent segment store.